Modelling of Clock Behaviour

Donald B. Percival*
Applied Physics Laboratory
Box 355640
University of Washington
Seattle, WA 98195–5640
USA

Abstract

Although modern atomic clocks can keep time with amazing accuracy, there are limits to their capabilities. These limitations can be assessed with the help of statistical models. We present a basic introduction to the ideas behind the modelling of clock behaviour and consider some opportunities for improving both clock modelling and its use for interpreting and making efficient use of clock data.

1 Introduction

Modern atomic clocks can keep time to accuracies that were unimaginable a hundred years ago. For systems such as Galileo that depend critically on precise time, there is a need to evaluate the performance of individual clocks and ensembles of clocks as accurately as possible. Any predictable deviations that an atomic clock makes away from 'perfect' time could in principle be corrected for to improve the accuracy of the clock; however, at some level of accuracy, all real clocks exhibit unpredictable deviations from perfect time. These deviations are random in nature, and the description of these deviations involves the use of statistical models. In Section 2 we give a basic introduction to the ideas behind statistical modelling of clock behaviour, with an emphasis on what kinds of information are encapsulated in these models. We describe in Section 3 a set of five canonical models that are commonly used as a starting point for modelling clocks. The final section (Section 4) is devoted to a brief discussion of some opportunities for improving the manner in which clocks are modelled and the manner in which the models are used in practical applications such as Galileo.

2 What is Clock Modelling?

Let's start with the following thought experiment. We have a clock whose ability to keep time we want to assess. We assume that, whenever we want, we can compare the time keep by this clock against some good approximation to 'perfect' time (in practice this would be defined as the time kept by a large ensemble of high precision clocks). Suppose we set our clock to agree exactly with perfect time at midnight on a certain day. We then observe how our clock deviates from perfect time over the next 24 hours. The upper panel of Figure 1 shows a plot of these hypothetical deviations versus elapsed time from midnight (measured in hours). We see that our clock wandered away from perfect time over the 24 hour period

*e-mail: dbp@apl.washington.edu, Phone: +1 206 543 1368, Fax: +1 206 543 6785

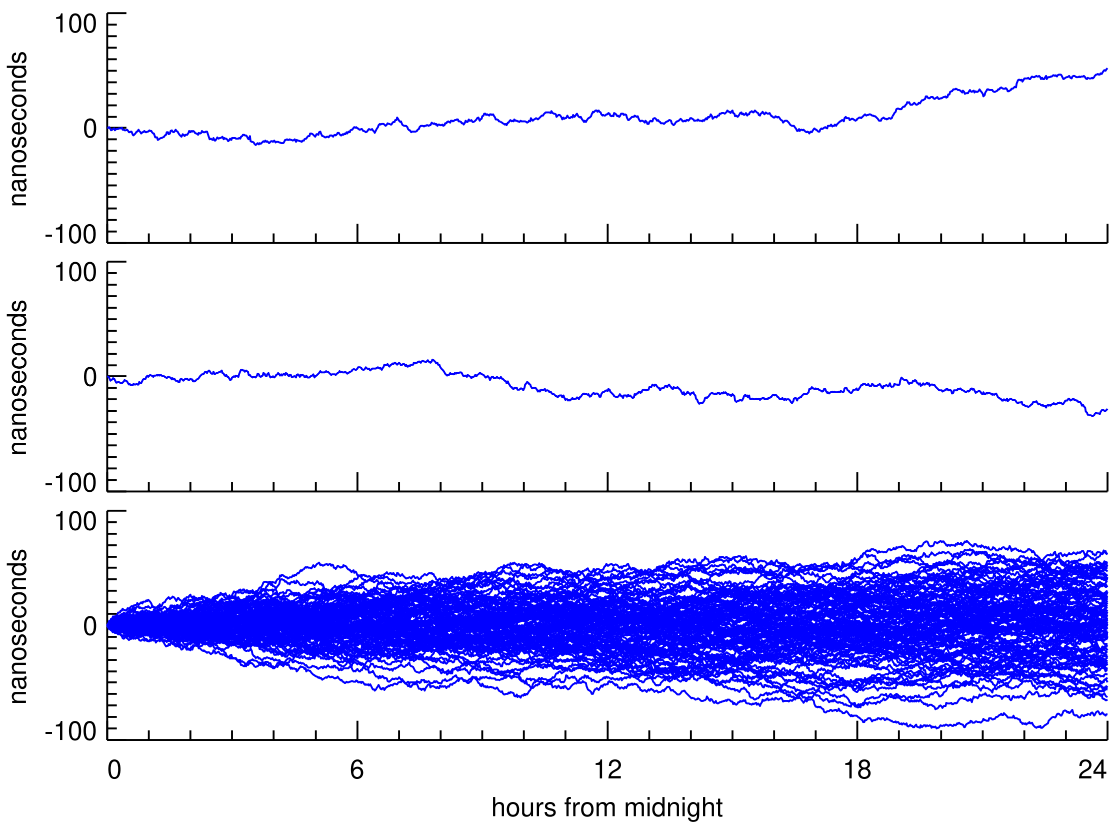

Figure 1: Simulated deviations between the time kept by a hypothetical clock and 'perfect' time over an elapsed period of a day. The clock is adjusted to agree with perfect time at the beginning of each day. The upper two plots show the time deviations for the first two days of this thought experiment, while the lower plot shows an overlay of the deviations for 100 consecutive days. A positive (negative) deviation means that the clock displayed a time that was ahead (behind) of what a clock keeping perfect time would display.

and, at the end of the day, had gained about 50 nanoseconds (i.e., a billionth of a second; the choice of nanoseconds as the unit of measure for the time difference is entirely arbitrary). Suppose we reset our clock to agree with perfect time again at midnight and record another set of measurements over the next 24 hours. These are shown in the middle panel of Figure 1. The deviations are not the same as what we saw before. The clock now ends up losing about 30 nanoseconds after 24 hours; however, visually the deviations in the upper and middle plots look similar in the sense that they both have a similar 'bumpiness' to them.

Suppose we repeat this experiment for 100 consecutive days. The bottom plot shows the 100 series of deviations that we observed plotted on top of each other. Each individual series tends to have the same bumpiness that we observed in the first two days of our experiment, but each series is unique. From this experiment, we learn that we cannot predict exactly the deviations for any given day, but the deviations do have similar characteristics from day to day. For example, the clock never seems to wander more than about 100 nanoseconds away from perfect time, but, on the other hand, it usually wanders at least 10 nanoseconds away within 6 hours of its being reset to perfect time. Under the assumption that the mechanism

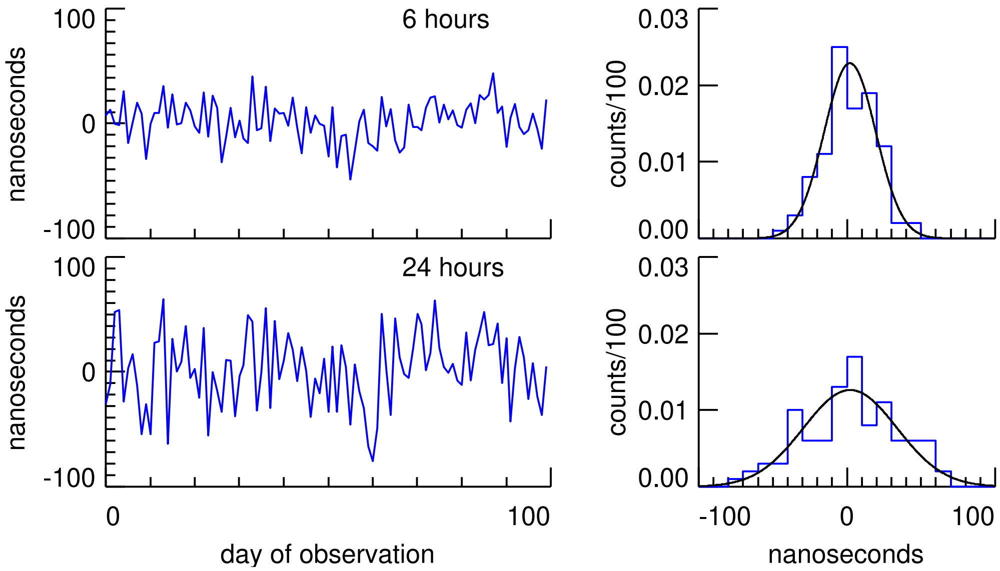

Figure 2: Simulated time deviations of a clock from 'perfect' time at two fixed elapsed times from midnight over a 100 day period, along with histograms for the deviations and a fitted Gaussian (normal) distribution. The upper left-hand plot shows the time deviations at an elapsed time of 6 hours (i.e., 6AM) for 100 consecutive days, while the upper right-hand plot shows the corresponding histogram (staircase curve) and fitted theoretical Gaussian distribution (bell-shaped curve). The bottom row shows the same plots for the time deviations at an elapsed time of 24 hours.

that is responsible for the observed deviations does not change from day to day, we can use our collection of measurements to make some statistical statements about the nature of the time deviations within a 24 hour period. For example, the average deviation after 6 hours is close to zero, with 95% of the individual curves ranging roughly between -30 and $+30$ nanoseconds. After 24 hours, the average deviation is still close to zero, but now 95% of the curves range roughly between -60 and $+60$ nanoseconds. By collecting more and more data, we can build up a complete statistical description of how our clock tends to deviate from perfect time over a period of a day.

The purpose of a clock model is to summarize the information collected in experiments like the above. As an example, let us focus again on the deviations that we observe at 6AM on each of the 100 days. These 100 deviations are plotted in the upper left-hand plot of Figure 2 versus the number of the day on which they were recorded, and a histogram for these deviations is given by the 'staircase' curve in the upper right-hand plot. The histogram offers a summary of the distribution of the 100 6AM deviations and shows that indeed the deviations are centered around zero and that most of them occur between -30 and 30 nanoseconds. We can further summarize the observed distribution by using a theoretical distribution that depends upon a small number of parameters. One theoretical distribution that has been used extensively is the Gaussian (or normal) distribution, which depends on two parameters, namely, the mean and the variance. The mean is determined by the average of the 100 6AM deviations and is close to zero. The variance is a measure of how spread out the deviations are and quantifies the fact that most of the values occur between -30 and

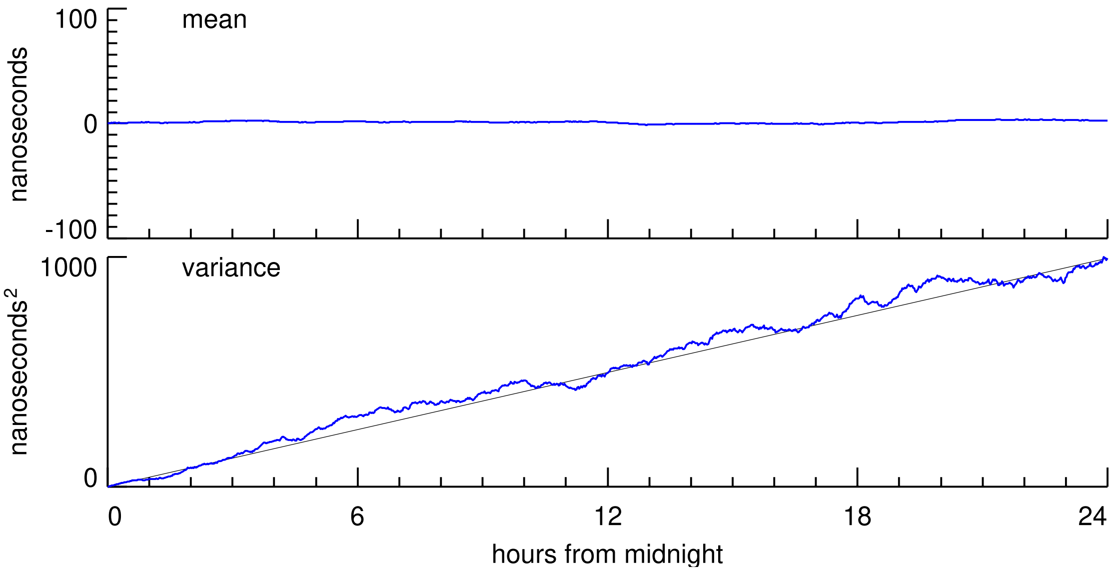

Figure 3: Sample means and variances for the 100 time deviation series shown in the bottom plot of Figure 1 versus elapsed time. The top plot indicates that the means are close to zero for all elapsed times, whereas the jagged curve in the bottom plot shows that the variances tend to be approximately proportional to elapsed time (the line on this plot is a linear approximation to the jagged curve).

30 nanoseconds. The larger the variance is, the more spread out the associated histogram would be. The fitted Gaussian distribution for the 100 6AM deviations is shown by the bell-shaped curve in the upper right-hand plot. Thus, if we were to contemplate continuing our experiment beyond 100 days, we cannot predict what value we would observe at 6AM on any given future day, but, under the assumption that what we will observe in the future will be similar to what we observed in the past, the simple two parameter model that we have derived can be used to answer questions about how likely any future 6AM measurement is to be, say, within a certain interval (e.g., -10 and 10 nanoseconds).

We can obviously use a similar procedure to model the deviations collected at elapsed times besides 6 hours after midnight. The bottom plots of Figure 2 show corresponding results for our hypothetical experiment after 24 hours. Note that the Gaussian distribution is more spread out than in the 6 hour case, which is reflected in the fact that the variance is larger. We can create similar models for any given elapsed time between 0 and 24 hours. Figure 3 shows the mean (upper plot) and variance (jagged curve in the lower plot) for the fitted Gaussian distributions as a function of elapsed time. We see that, while the mean is always close to zero, the variance tends to increase with time in a linear manner. This result opens up the possibility of summarizing the means and variances over all elapsed times between 0 and 24 hours. The summary consists of a Gaussian distribution associated with each elapsed time, with a mean of zero for each distribution and a variance that is proportional to the elapsed time; i.e., we have boiled down the means and variances to a single parameter, namely, the constant of proportionality that sets the actual value of the variance (this parameter is essentially the slope of the line shown in the bottom plot).

The Gaussian distributions that we have considered so far are all univariate distributions;

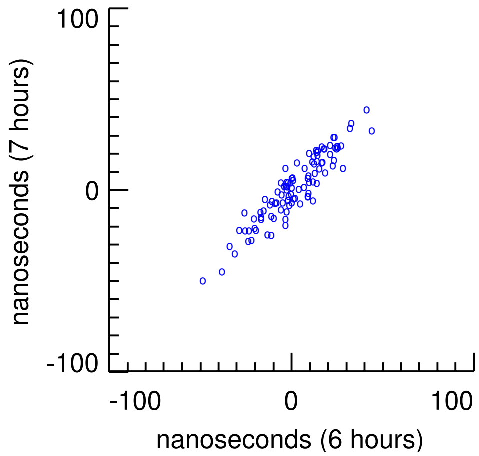

Figure 4: Plot of the observed time deviation at 7AM versus the corresponding deviation at 6AM for each of the 100 days over which we conducted the thought experiment recorded in Figure 1. The two time deviations tend to be positively correlated; i.e., a large positive or negative deviation at 6AM tends to followed by a similar deviation at 7AM.

i.e., they can tell us something about where we might expect future deviations to be at a given elapsed time, but they don't tell us anything about how observed deviations at distinct elapsed time are related to one another. Figure 4 is a first look at the multivariate properties of our experiment. Here we have plotted the time deviation observed at 7AM versus the deviation at 6AM for all 100 such pairs of data. We see that a large positive or negative value at 6AM tends to be associated with a simular value at 7PM; i.e., the time deviations at these two elapsed time are positively correlated. In addition to univariate properties, clock models are designed to summarize bivariate properties such as the positive correlation indicated in Figure 4 and also corresponding higher order multivariate properties involving the statistical relationship between the time deviations at three or more distinct elapsed time.

3 Five Canonical Clocks Models

Since the 1960s, the standard way of formulating a multivariate clock model that can usefully summarize all the salient statistical properties of data similar to that in our thought experiment is via a canonical set of five so-called 'power law' models. The five models are classified by a parameter often denoted by the symbol α (this is actually an exponent that determines the manner in which a theoretical quantity known as the power spectral density function decays as a function of a variable known as Fourier frequency). For the canonical models the parameter α can assume one of 5 values, namely, 0, -1, -2, -3 and -4. The corresponding models are known as, respectively, white phase noise ($\alpha = 0$), flicker phase noise ($\alpha = -1$), random walk phase noise ($\alpha = -2$), flicker frequency noise ($\alpha = -3$) and random walk frequency noise ($\alpha = -4$). Each of the five models in addition depends upon one additional parameter, which we will denote as C and refer to as the level parameter. Specification of a canonical clock model consists of determining which of the 5 values of the power law parameter α is most appropriate and then setting the level parameter C.

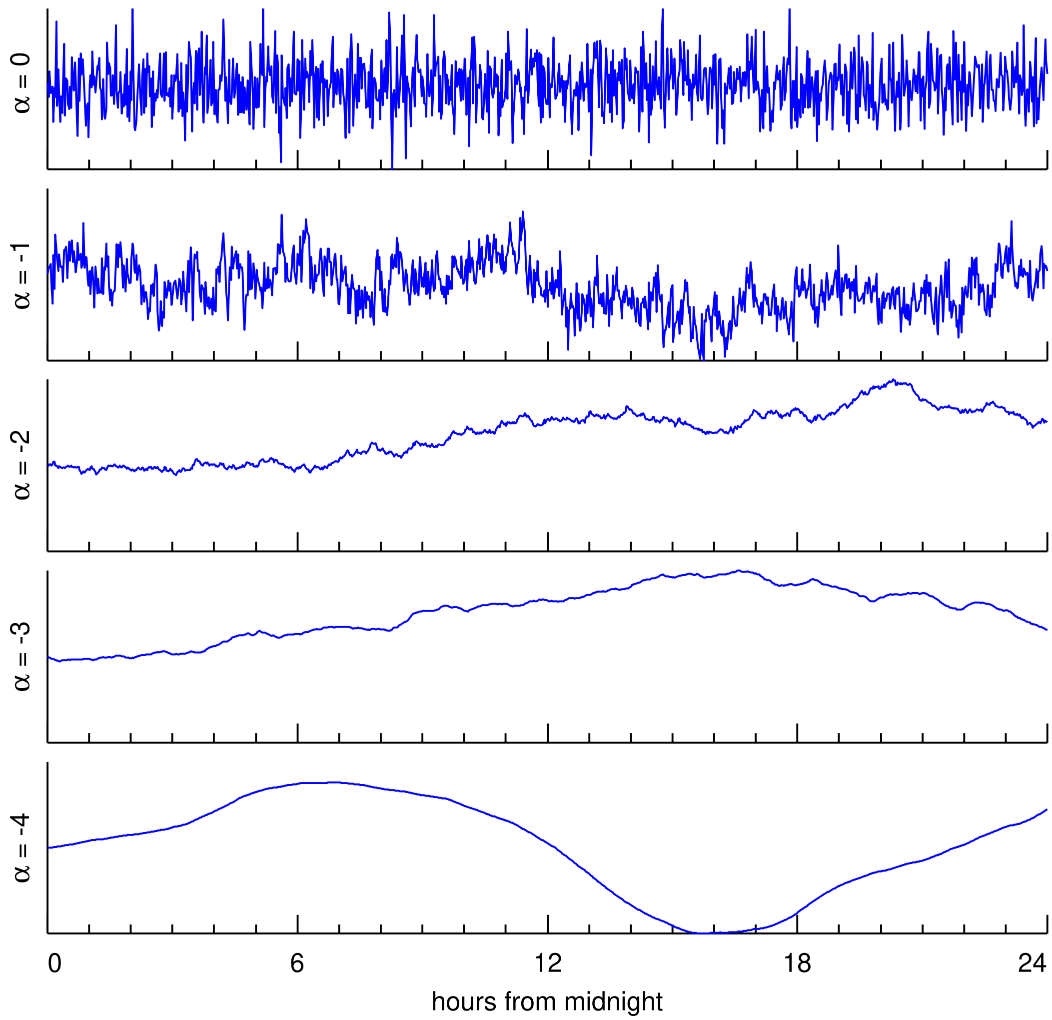

Figure 5: Like the top plot of Figure 1, but now for a hypothetical clock well modelled by, from top to bottom, white phase noise ($\alpha = 0$), flicker phase noise ($\alpha = -1$), random walk phase noise ($\alpha = -2$), flicker frequency noise ($\alpha = -3$) and random walk frequency noise ($\alpha = -4$).

The clock model behind our thought experiment is actually random walk phase noise ($\alpha = -2$), so the time deviations in Figure 1 can be regarded as examples of what we can expect to see from a clock that is well-modelled by such noise. Figure 5 shows one example of what we might see in one replication of our thought experiment in which the clock under assessment is well-modelled by one of the five canonical models. Visually, these five models yield time deviations, each of whose characteristic 'bumpiness' is rather distinct from the other four. The deviations are very erratic looking for $\alpha = 0$ and -1 and are much smoother for $\alpha = -3$ and -4.

The usual formulation of the canonical models is based on a multivariate version of the Gaussian distribution. The associated univariate distributions all have means of zero at each elapsed time, but the choice of α sets the way in which the variance changes over time. When $\alpha = -2$, we saw from Figure 3 that the variance grows linearly with elapsed time;

for $\alpha = 0$, the variance is a constant for all elapsed time; and, for $\alpha = -4$, the variance grows quadratically (for $\alpha = -1$ and $\alpha = -3$, the variance grows with elapsed time, but in a manner that is not easy to describe). The level parameter C just adjusts these variance curves to have the proper values. In effect, C does nothing more than to specify the labelling on the vertical axes in plots like those in Figure 1. It does not influence the characteristic 'bumpiness' associated with a particular α, but does determine whether or not the deviations after 24 hours span, e.g., a few microseconds, tens of nanoseconds or a couple of picoseconds.

The choice of a canonical model for a clock consists of setting values for α and C. Once these two values are chosen, if the model is a good one for a particular clock, we have all the information that we need to assess the statistical properties of its time deviations. This information can be used to assess how well a particular clock is doing relative to other clocks. In particular, if we have modelled an entire ensemble of clocks, we can use these models to determine how best to combine them together to generate an 'ensemble time' that is more accurate than the time generated by any one clock – all this from knowing just two values for each clock!

4 Opportunities for Improving Clocks Models and Their Usage

In practice, the picture that we have painted so far of clock modelling has been too simplistic. There are at least four 'real-world' complications that make it harder to reap the potential benefits of clock modelling.

1. It has long been recognized that a good choice for a canonical model for an actual clock can depend upon the elapsed time; i.e., a model that is appropriate for a maximum elapsed time of a day is usually not also appropriate for maximum elapsed times of a second or a year. To get around this problem, the common practice has been to entertain possibly different canonical models for different elapsed times for a given clock. In practice, two numbers are usually not enough to adequately model an actual clock.

2. There are some clocks whose time deviations are not well described by any one of the five canonical 'bumpiness' curves in Figure 5. There are no compelling physical arguments to support use of these particular five models, and their prevalence is somewhat of an historical artifact. In fact, there is a well-defined multivariate Gaussian model associated with all possible values for α, and the ones for which α is between, say, -1 and -2 would generate time deviations whose 'bumpiness' offers a smooth transition between the second and third plots in Figure 5; i.e., there is a big qualitative jump between flicker phase noise ($\alpha = -1$) and random walk phase noise ($\alpha = -2$) that can be bridged if we were to entertain α's between these two numbers.

3. In our thought experiment, we were able to collect enough data so that we could reliably make statements about the statistical properties of our hypothetical clock over the span of a day. If we increase the maximum elapsed time to a year, we could never collect a similar amount of data (i.e., 100 years worth) for any actual clock. Faced with a lack of data, determination of model parameters becomes a very challenging problem.

4. Our thought experiment concentrated on the situation where the time deviations were 'truly random' in the sense that there was no tendency for our theoretical clock to consistently gain or lose time over a given elapsed time; i.e., the average of the deviations was close to zero, as demonstrated in the top plot of Figure 3. In practice, this is not always the case. Any average deviation that is not close to zero is an indication of a 'trend' in the clock. Trend is of interest because, if properly determined, we can enhance the performance of any clock by merely adjusting its displayed time by subtracting off an amount dictated by the trend. Usually trend must be estimated from the data, and this becomes a particularly challenging problem when we are faced with a limited amount of data.

In view of the above complications, there are a number of ways in which the modelling of clocks can be improved upon to bring it more into line with modern statistical practice. The first has to do with limiting ourselves to just five canonical models. Rather than insisting upon setting α to a canonical value, we can estimate it from a continuum of values. Explicit estimation of α would help alleviate a statistical problem with the common practice of looking at a series of time deviations and selecting a canonical α by some means, after which the analysis continues under the assumption that α was known in advance; i.e., it is assumed that we do not need to consider the effect of sampling variability on the selection of α. The confidence bounds that are attached to estimates of common clock stability measures (e.g., the Allan variance) invariably are based on the notion that α is known and that all sampling variability is due to uncertainty about the level parameter C. When we have a limited amount of data, determination of α becomes problematic and is prone to the vagaries of sampling errors, but current practice is to ignore this potentially important source of variability.

In a similar manner, estimation of trends in time deviations is usually not handled in a manner such that the uncertainty in the estimated trend is propagated through to other quantities; i.e., the detrended time deviations are often treated as if they were the actual measurements, and clock stability measures are estimated from them. An assessment of uncertainty in these estimated measures does not take into account uncertainties in estimating the trend. While treating α and/or the trend component as if they were not subject to sampling errors is convenient from the point of view of computations, it is certainly questionable from a statistical point of view.

Finally let us offer some speculations about the use of clock models in the generation of time scales from ensembles of clocks. An elegant mathematical approach for forming such scales is based upon the Kalman filter. The usual Kalman filter, however, makes the implicit assumption that a model for particular clock is known *perfectly*; i.e., any sampling variability due to only having a finite amount of data to determine α and C is ignored by the Kalman filter. Even though the Kalman filter is known as an 'optimal' filtering procedure, it is only so under the unrealistic assumption of perfectly known clock models. Amongst those who have studied the formation of time scales, the Kalman filter is regarded as problematic and inferior to certain *ad hoc* procedures that appear to form better time scales than this so-called 'optimal' procedure. It is our contention that this dissatisfaction with the Kalman filter is well-founded and is tied to the fact that uncertainty in estimated parameters is not being handled properly. This deficiency is particularly evident in an operational environment in which clocks are added to a time scale after a relatively short calibration period, during which models are determined with a paucity of data and hence have parameter estimates with necessarily nonnegligible sampling errors. What seems to be called for is a reformulation of the Kalman filter that takes into account this important source of uncertainty. Thus clocks with seemingly very nice statistical properties would be downweighted appropriately if there is large uncertainty in their estimated parameters.